Cormen 5.4: uses of indicator variables

5.4.1 Birthday paradox (slight generalization)

We have k persons holding some number between 1 and n

How large should k be before we expect two to have the same number?

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ hold the same number} \\ 0 & \text{else} \end{cases} \qquad i,j \in \{1,2,\dots,k\}$$

$X = \sum_{1 \le i < j \le k} X_{ij}$ is the number of pairs of persons who hold the same number

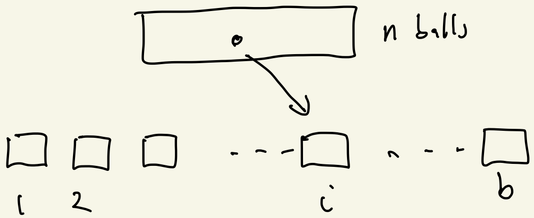
$$E(X) = E\Big(\sum_{1 \le i < j \le k} X_{ij}\Big) = \sum_{1 \le i < j \le k} E(X_{ij}) = \sum_{1 \le i < j \le k} \frac{1}{n} = \binom{k}{2} \cdot \frac{1}{n}$$

So $E(X) = \dfrac{k(k-1)}{2n} \ge 1$ when $k \ge \sqrt{2n} + 1$

if $n = 366$ we set $k \ge 28$

$$\frac{28 \cdot 27}{2 \cdot 366} \sim 1.03$$

5.4.2 Balls in bins (kuler i kasser)



$p(\text{given ball} \to \text{bin } i) = \frac{1}{b} \quad \forall i$

$$X_{ij} = \begin{cases} 1 & \text{if ball } i \text{ lands in bin } j \\ 0 & \text{else} \end{cases} \qquad p(X_{ij} = 1) = \frac{1}{b}$$

$$X_j = \sum_{i=1}^{n} X_{ij} \quad \text{\# balls in bin } j$$

$$E(X_j) = E\left(\sum_{i=1}^{n} X_{ij}\right) = \sum_{i=1}^{n} E(X_{ij}) = \sum_{i=1}^{n} \frac{1}{b} = \frac{n}{b}$$

Expected # of balls to throw before there is one in a given bin j is $\frac{1}{1/b} = b$

E(# balls before all bins are nonempty) $= O(b \log b)$

(same analysis as for coupon-collector)

phases $0, 1, 2 \ldots n$ in phase i there are i non-empty bins phase

Expected no balls to progress

0	$p = \frac{b}{b} = 1$	1
1	$p = \frac{b-1}{b}$	$\frac{b}{b-1}$
2		
i	$p = \frac{b-i}{b}$	$\frac{b}{b-i}$
$i+1$		
$\vdots$		
	$p = \frac{1}{b}$	b

So Expected no of balls before all non-empty

$$= \sum_{i=1}^{b} \frac{b}{i} = b \sum \frac{1}{i} = O(b \ln b)$$

5.4.3 Streaks

Flip a fair coin n times. Outcome h/t each time

A streak is a sequence of flips with the same value:

....thhh..ht...   or   ....htt..th...

Let L be the longest streak of heads when flipping a fair coin n times

A_{ik} : only heads in flips $i, i+1, .., i+k-1$

$p(A_{ik}) = 2^{-k}$

so for $k = 2\lceil \log n \rceil$ we get

$$p(A_{i\,2\lceil \log n\rceil}) = 2^{-2\lceil \log n \rceil} \leq 2^{-2\log n} = n^{-2}$$

There are at most $n - 2\lceil \log n \rceil + 1$ positions when a streak of heads of length $2\lceil \log n \rceil$ can start so

$$(*) \quad p\left(\bigcup_{i=1}^{n-2\lceil \log n\rceil+1} A_{i,2\lceil \log n \rceil}\right) \leq \sum_{i=1}^{n-2\lceil \log n \rceil+1} \frac{1}{n^2} \quad \text{by union bound}$$

$$< n \cdot \frac{1}{n^2} = \frac{1}{n}$$

What is the expected length of a longest streak of heads?

L_j : event that longest streak has length j ($L_i \cap L_j = \emptyset$ $i \neq j$)

L = length of longest streak

$$E(L) = \sum_{j=0}^{n} j \cdot p(L_j) \quad \text{by definition of } E(\ )$$

$$p(\text{streak of length} \geq 2\lceil \log n \rceil \text{ anywhere})$$

$$\leq p\left(\bigcup_{i=1}^{n-2\lceil \log n \rceil+1} A_{i,2\lceil \log n \rceil}\right) < \frac{1}{n} \quad \text{by } \circledast!$$

$$E(L) = \sum_{j=0}^{n} j \cdot P(L_j)$$

$$= \sum_{j=0}^{2\lceil \log n \rceil-1} j \cdot p(L_j) + \sum_{j=2\lceil \log n \rceil}^{n} j \cdot p(L_j)$$

$$< 2\lceil \log n \rceil \sum_{j=0}^{2\lceil \log n \rceil-1} p(L_j) + n \sum_{j=2\lceil \log n \rceil}^{n} p(L_j)$$

$$< 2\lceil \log n \rceil \cdot 1 + n \cdot \frac{1}{n}$$

$$= O(\log n)$$